

Performance Evaluation for Particle Filters

Rémi Chou
Supélec, Metz Campus, France
Georgia Tech Lorraine, France

Yvo Boers, Martin Podt
Thales Nederland B.V., The Netherlands

Matthieu Geist
IMS Research Group
Supélec, Metz Campus, France

Abstract—Performance evaluation in particle filtering problems is commonly performed via point estimator comparison. However, in non-Gaussian cases, this can be not always meaningful and entire particle clouds need to be compared. The Kullback-Leibler divergence (KLD) can be used for such a particle cloud comparison. In contrast to KLD estimates commonly used in particle filtering applications, we present an estimator of the KLD being applicable to any cloud of particles. This estimator is applied to a performance evaluation scheme generally relevant to any particle filter, of which abilities are equal to no other known scheme in the literature. Through simulations and concrete examples, we will show that it is suitable to practically compare particle clouds, which have a limited number of particles, have a different size, are close to each other and have an high dimensionality.

Keywords: Particle filter, particle filter comparison, Kullback-Leibler divergence, nearest neighbor.

I. INTRODUCTION

Many real life problems involve estimating unknown dynamic quantities on the basis of observed data. Some examples are : tracking a target on the basis of radar measurements, estimating a signal in a noisy communication channel, GPS navigation or estimating the volatility of financial indices on the basis of stock market data. Particle filters [1] are commonly used for such nonlinear filtering problem. In order to find the most suitable and effective filter for a given problem, convergence analysis and comparisons of particles filters are needed. The purpose of this paper is to present a practical and generally applicable tool to do so.

Particle filter performance analysis is generally done using the mean square error between point estimators. If it is fine for Gaussian cases, it is not for problems such that high order moments cannot be well approximated by a Gaussian density, as demonstrated in [2]. In that case, another measure has to be used and the comparison of entire posterior distributions is needed. One of the numerous possibilities to compare two probability density functions (pdfs) is the Kullback-Leibler Divergence (KLD).

A special interest has been shown toward the KLD, that has demonstrated its use for measuring the difference between two distributions in a wide range of applications (speech recognition [3], image registration [4], classification scheme [5], etc.). This interest is probably raised by its intuitive understanding related to information theory (see [6]).

Therefore, it is not surprising to also find the KLD in particle filter applications, like in [7] to adapt the number of particles over time, or in [8] [9] to evaluate convergence performance of different particle filters.

However, a major issue is to estimate the KLD for densities only known and described by a limited number of samples, which might not share the same support, like in particle filtering applications. Moreover, even if a natural idea is to use histograms, we will see that such a method is not applicable to a state vector with high dimensionality (practically greater than 2 or 3). Usually, only non generalizable methods are used to evaluate the KLD in very specific cases. For example, the KLD is computed for Gaussians in [3], for Gaussian mixtures in [4], for one dimensional histograms in [5], for regularized particle filters in [8], for samples that have the same size and the same support in [9]. To the best of our knowledge, no generally applicable KLD estimate in particle filter applications can be found in the literature.

The contributions of this paper are the use of the generally applicable KLD estimate, presented in [10], in a particle filtering context, and its application to a performance evaluation scheme applicable to any particle filter. The practical characteristics of this scheme concerning the following points will be demonstrated:

- ability to deal with a limited number of particles;
- ability to deal with clouds of different size;
- suitability to high dimension;
- ability to discriminate close distributions;
- interchangeability, under some restrictions.

The rest of the paper is organized as follows. In Section 2, the definition of the KLD is reviewed, and the chosen estimator presented. In Section 3, we demonstrate through simulations and concrete examples the practical abilities of the proposed estimator.

II. KULLBACK-LEIBLER DIVERGENCE

In this Section, let p and q be two continuous densities on \mathbb{R}^d .

A. Definition

The Kullback-Leibler divergence [6] $D_{KL}(p, q)$, between p and q is defined as the expected value of the log likelihood

ratio between p and q :

$$D_{KL}(p, q) := \mathbf{E}_p \left[\log \frac{p}{q} \right] = \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} \quad (1)$$

when the support of q is included in p one's, otherwise $D_{KL}(p, q) := +\infty$.

B. Properties

- $\forall p, q, D_{KL}(p, q) \geq 0$;
- $\forall p, q, D_{KL}(p, q) = 0 \Leftrightarrow p = q$.
Furthermore, a small value for $D_{KL}(p, q)$ indicates that p and q are close to each other;
- The KLD is not a true distance, since it is not symmetric in general and does not satisfy the triangle inequality. However, it is still appropriate as a benchmark to evaluate the closeness of a density to another.

Remark: the first argument of D_{KL} is the reference, whereas the second is the term to be compared to the reference. This can be understood thanks to the “coding penalty” interpretation in information theory, when a given probability density function is approximated by another (see [6] for further details).

C. Estimator [10]

Let $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_m\}$ be i.i.d samples drawn independently from p and q , respectively. In [10], an asymptotical unbiased and mean-square consistent estimator $\hat{D}_{KL}(p, q)$ of $D_{KL}(p, q)$, based on the k -Nearest Neighbor (NN) density estimation [11], is defined as follows

$$\hat{D}_{KL}(p, q) =$$

$$\frac{d}{n} \sum_{i=1}^n \log \frac{\nu_{k_i}(i)}{\rho_{l_i}(i)} + \frac{1}{n} \sum_{i=1}^n [\psi(l_i) - \psi(k_i)] + \log \frac{m}{n-1} \quad (2)$$

where $\nu_{k_i}(i)$ is the Euclidian distance¹ from X_i to its k_i -NN in $\{Y_j\}$, $\rho_{l_i}(i)$ the Euclidian distance between X_i and its l_i -NN in $\{X_j\}_{j \neq i}$, and ψ is the Digamma function, defined as the logarithmic derivative of the Gamma function.

Remark: small values for the k_i and l_i in (2) lead to a small bias but an high variance of the estimator. Here, the variance will be easily decreased by an average of several runs of the estimator, thus small k_i and l_i should be chosen. Nevertheless, if the number of samples is large enough, the k_i and l_i can be selected larger, to reduce the variance while still guaranteeing a small bias.

III. COMPARISON OF CLOUDS

Our goal is to compare two clouds $\{(\mathbf{x}_i, w_i^x)\}_{i \in \{1, 2, \dots, N\}}$ and $\{(\mathbf{y}_j, w_j^y)\}_{j \in \{1, 2, \dots, M\}}$, obtained from any particle filters. It is assumed that $\forall i \in \{1, 2, \dots, N\}, \forall k \in \{1, 2, \dots, M\}, w_i^x = 1/N$ and $w_j^y = 1/M$ (resampling² might be necessary). Because of this, some particles of the

clouds might have the same value. In the estimator of the KLD, this is taken into account by taking the k -Nearest Different Neighbor (NDN): if the k -NN of a point x is itself, then k is incremented until x and its k -NN are different. Consequently the l_i (respectively the k_i) in equation (2) are not necessarily equal among themselves. In the following simulations, the 1-NDN is always chosen.

A. Convergence

The estimator presented in Section II-C is asymptotically unbiased and mean-square consistent under the assumption of i.i.d data samples and some regularity properties (see [10]). The purpose of this Section is to show the convergence abilities of the estimator in practical particle filter applications, that is, when only a limited number of non i.i.d samples is available.

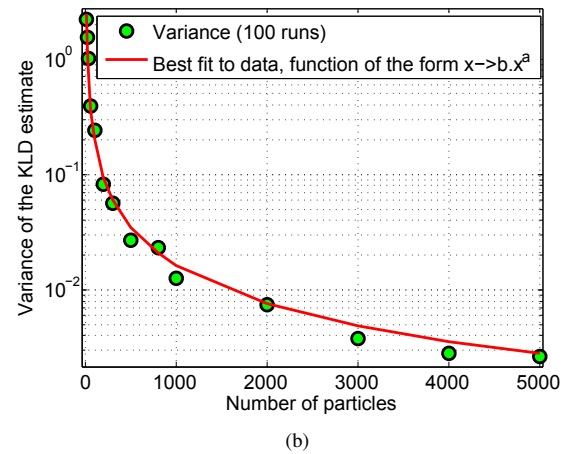
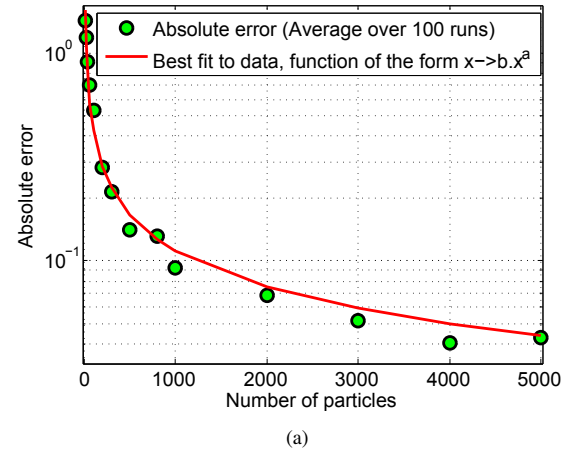


Figure 1. Absolute error (a) and Variance of the KLD estimate (b) vs. number of particles.

The following model is considered :

$$\begin{cases} s_{t+1} &= s_t + w_t \\ z_t &= s_t + v_t \end{cases} \quad (3)$$

¹Other distances can also be used, see [10].

²The “systematic resampling” algorithm in [1] can for instance be used.

where $w \sim \frac{1}{2}\mathcal{N}(-1, 0.1^2) + \frac{1}{2}\mathcal{N}(1, 0.1^2)$, $v \sim \mathcal{N}(0, 2^2)$, $s_0 \sim \mathcal{N}(0, 1)$.

The simulation consists in comparing two clouds C_1 , C_2 of identical size ($|C_1| = |C_2|$), obtained at time step $t = 10$, for two runs of the same Sampling Importance Resampling (SIR) filter [1]. A number of particles between 100 and 5000 will be used for $|C_1| = |C_2|$. Note that the true KLD between C_1 and C_2 is 0. Only the nearest different neighbor in the KLD estimation is taken into account. Moreover, the best (mean-square error) fit to data for a function of the form $x \mapsto bx^{-a}$ is computed.

Simulation results:

- **Bias:** for an average over 100 runs, $a = -0.58$ and $b = 6.15$ (see figure 1(a)) are obtained, that is, the error decreases approximately like $\frac{1}{\sqrt{N}}$.
- **Variance:** $a = -1.09$ and $b = 30.25$ are obtained (see figure 1(b)), that is, the variance decreases approximately like $\frac{1}{N}$.

The relative small values obtained for both bias and variance make the use of the estimator possible in practical cases, even if the particles are non i.i.d and in limited number.

Remark: these bias and variance should be ensured to be small enough in the case of close distributions comparison. This issue will be addressed in Section III-D.

Example : given the previous results concerning the order of magnitude of bias and variance of the KLD estimate, we can already now present a concrete example, which puts into exergue the fact that in contrast to the KLD, the mean-square error cannot always perform a meaningful particle cloud comparison.

Model (3) is considered with the same setup as previously, and the following filters are considered : a SIR filter \mathcal{PF}_{ref} with 10^6 particles (its clouds are consider as the “ground truth”), another SIR filter \mathcal{PF}_{1000} with 1000 particles and a Kalman filter, for which v is considered as Gaussian instead of bimodal.

At each time step k , the cloud from \mathcal{PF}_{1000} , and 1000 samples drawn from the pdf $p(s_k|Z_k)$ obtained by the Kalman filter, are compared to the reference cloud, obtained by \mathcal{PF}_{ref} . In figure (2), the comparison is done via the KLD on the one hand and via the mean error on the other hand. Here is an illustration of the limitations of a point estimate error : the Gaussian density obtain by the Kalman filter approximates sufficiently well the second order moment, but not the higher order moments in general, see figure (3).

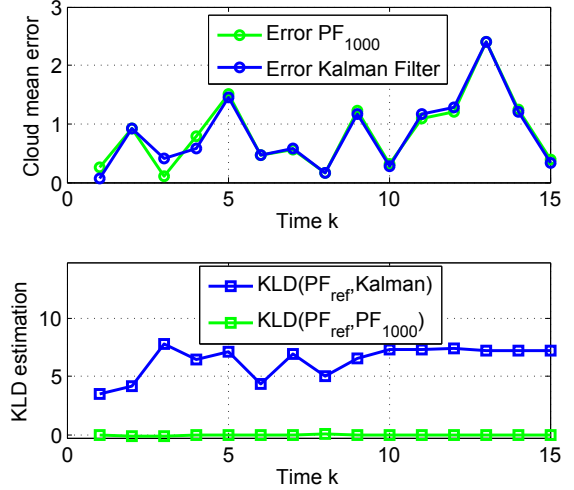


Figure 2. Estimation error for the mean of the clouds vs. KLD.

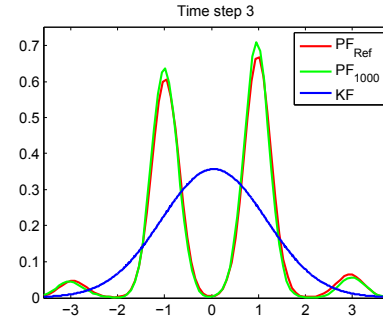


Figure 3. Representative pdf estimated by each of the three filters. The density plotted for \mathcal{PF}_{ref} and \mathcal{PF}_{1000} have been obtained by kernel smoothing.

B. Different Sample size

This Section evaluates the performance of the estimator for clouds of different size. This case can, for instance, occur in convergence analysis, where two clouds of different size obtained from the same filter are compared.

The model described in the previous Section by (3) is used in the following simulation, which consists in comparing two clouds C_1 , C_2 obtained at time step $t = 10$, for the run of two SIR Filters \mathcal{PF}_1 and \mathcal{PF}_2 . The number of particles of C_1 is fixed at 5000, whereas values between 100 and 5000 will be used for $|C_2|$. Note that the true KLD between C_1 and C_2 is 0. Furthermore, the results are compared with the ones obtained for $|C_1| = |C_2|$.

Simulation results: for both bias and variance (see figure (4)), the values observed in the case $|C_1| \neq |C_2|$ are comparable to the ones obtained in the case $|C_1| = |C_2|$, which shows the ability of the estimator to deal with clouds of different size.

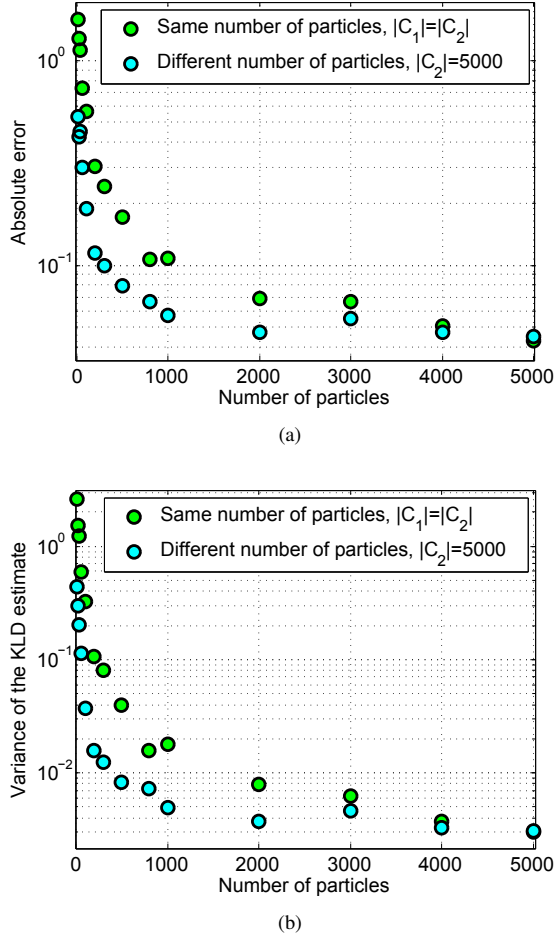


Figure 4. Absolute error (a) and Variance of the KLD estimate (b) vs. number of particles.

C. High dimension suitability

A natural idea to estimate $D_{KL}(p, q)$, is to use histograms with equally sized bins to estimate the densities p and q . However, histogram based methods suffer [11] among others from the “curse of the dimensionality”³, and are thus often unsuitable, especially if the dimension is higher than 2.

In order to demonstrate the ability of the proposed estimator to handle high dimensions, we consider a 3D target tracking problem for a target flying at a constant speed.

The state vector $\mathbf{s}_t = [x_t \ y_t \ z_t \ \dot{x}_t \ \dot{y}_t \ \dot{z}_t]^T$, consists of the target cartesian position and speed in kilometers. The system is described as follows:

$$\begin{cases} \mathbf{s}_{t+1} &= f(\mathbf{s}_t) + g(\mathbf{w}_t) \\ \mathbf{z}_t &= h(\mathbf{s}_t) + \mathbf{v}_t \end{cases} \quad (4)$$

³Since the number of bins grows exponentially with the number of dimensions, a very large number of samples is required in high dimensions, or else most of the bins would be empty.

with

$$f(\mathbf{s}_t) = \begin{bmatrix} 1 & 0 & 0 & T & 0 & 0 \\ 0 & 1 & 0 & 0 & T & 0 \\ 0 & 0 & 1 & 0 & 0 & T \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{s}_t, \quad (5)$$

$$g(\mathbf{w}_t) = \begin{bmatrix} \frac{1}{6}a_x^{max}T^2 & 0 & 0 \\ 0 & \frac{1}{6}a_y^{max}T^2 & 0 \\ 0 & 0 & \frac{1}{6}a_z^{max}T^2 \\ \frac{1}{3}a_x^{max}T & 0 & 0 \\ 0 & \frac{1}{3}a_y^{max}T & 0 \\ 0 & 0 & \frac{1}{3}a_z^{max}T \end{bmatrix} \mathbf{w}_t, \quad (6)$$

$$h(\mathbf{s}_t) = \begin{bmatrix} \sqrt{x^2 + y^2 + z^2} \\ \arctan(y/x) \\ \arctan(z/\sqrt{x^2 + y^2}) \\ (x.v_x + y.v_y + z.v_z)/\sqrt{x^2 + y^2 + z^2} \end{bmatrix}, \quad (7)$$

$\mathbf{w} \sim \mathcal{N}(0, 0.1^2)$ and $\mathbf{v} \sim \mathcal{N}(0, 0.1^2)$. The update time T is taken equal to 1s, the maximum accelerations a_x^{max} , a_y^{max} , a_z^{max} along x , y and z are taken equal to 5m/s, and the state vector is initialized with $\mathbf{s}_0 = [7 \ 5 \ 1 \ -0.2 \ 0.1 \ 0]^T$.

The simulations done are the same as in section III-A: two clouds C_1 , C_2 of identical size ($|C_1| = |C_2|$), obtained at time step $t = 5$, for two runs of the same SIR filter are compared. Values between 100 and 5000 are used for $|C_1| = |C_2|$. Note that the true KLD between C_1 and C_2 is 0. Only the nearest different neighbor in the KLD estimation is taken into account. Moreover, the best (mean-square error) fit to data for a function of the form $x \mapsto bx^{-a}$ is computed.

Simulation results:

- **Bias:** for an average over 100 runs, $a = -0.34$ and $b = 2.40$ (see figure 5(a)) are obtained.
- **Variance:** $a = -1.48$ and $b = 97.56$ are obtained (see figure 5(b)).

Even if the bias and variance are slightly larger than in the one dimensional case of section III-A, the relative small values obtained for both bias and variance, make the practical use of the estimator still possible in particle filter applications with high dimensionality.

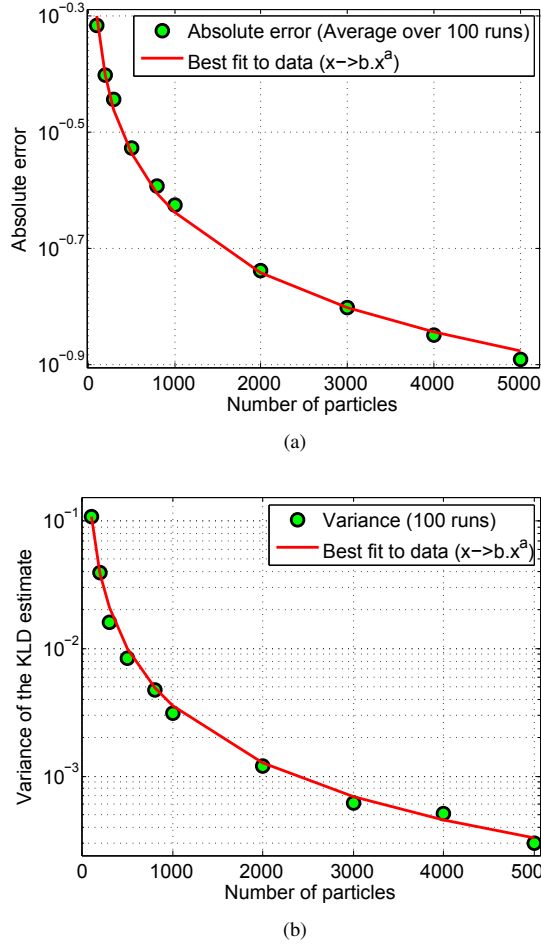


Figure 5. Absolute error (a) and Variance of the KLD estimate (b) vs. number of particles.

D. Discrimination of close distributions

As mentioned in section III-A, the estimator ability to compare close clouds has still to be ensured. The aim of this section is to show how to discriminate close clouds with the estimator \hat{D}_{KL} . A simple way to proceed is to average \hat{D}_{KL} in order to decrease its variance.

Let $\hat{D}_{KL}^{(n_0)}$ denotes an average of \hat{D}_{KL} over n_0 runs, the variance of \hat{D}_{KL} is thus expected to be reduced by a factor n_i . A simple example of close clouds discrimination is first considered: the KLD between a Normal and a Student distribution with parameter ν , both known by 1000 i.i.d samples, is computed⁴. A Normal distribution can be discriminated⁵ from a Student with $\nu = 11$ with $\hat{D}_{KL}^{(100)}$, whereas only a Normal from a Student with $\nu = 1$ can be discriminated with \hat{D}_{KL} , see figure 6. A comparison of the pdfs of a Normal and a Student distribution with $\nu = 1$ and $\nu = 11$ is shown on

⁴ $k = 10$ nearest neighbors in the estimation are taken into account, and the 3σ bounds for \hat{D}_{KL} and $\hat{D}_{KL}^{(100)}$ are computed with 100 runs.

⁵The 3σ bounds of the KLD between the Normal and the Student(ν), and the 3σ bounds of the KLD between two Normals, do not overlap.

figure (7).

This example shows the ability of \hat{D}_{KL} to discriminate close distributions. Consider now a more concrete example :

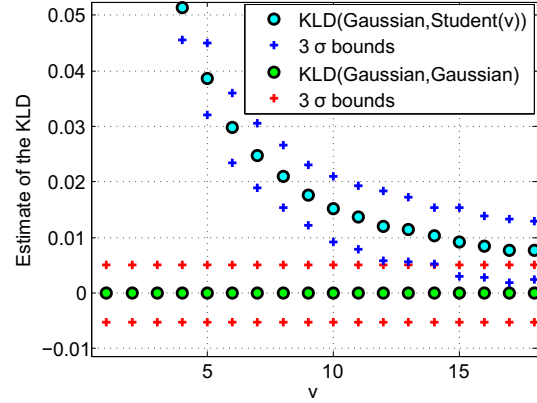


Figure 6. Discrimination of a Normal and a Student distribution with parameter $\nu = 11$ with $\hat{D}_{KL}^{(n_i)}$.

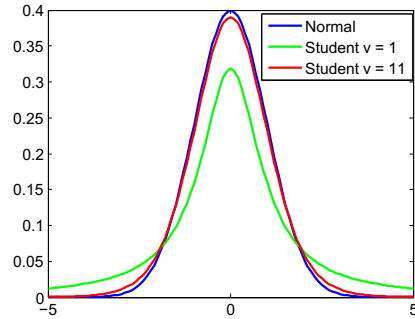


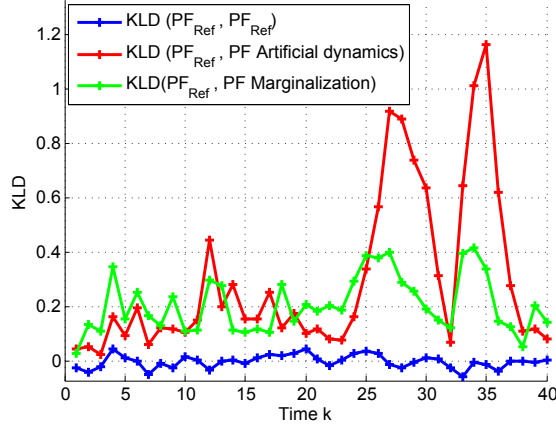
Figure 7. Normal and Student distributions with parameter $\nu = 1$ and 11.

Assume a model with an unknown parameter, and two filters to be compared. The model is

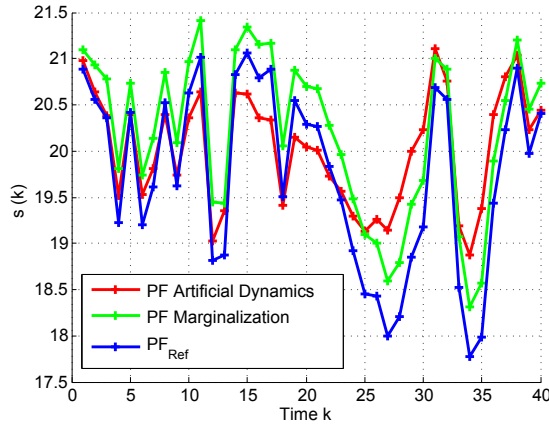
$$\begin{cases} s_{t+1} &= \frac{s_t}{2} + 10 + w_t \\ z_t &= s_t + \theta + v_t \end{cases} \quad (8)$$

with θ a parameter in $[10, 12]$. $\theta = 11.7$ is taken for data generation. The filters to be compared are :

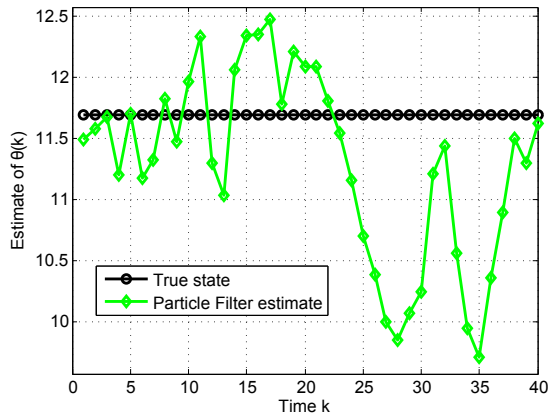
- a SIR filter, PF_{AD} , using 2000 particles, implementing an artificial dynamics method [12] to estimate θ , i.e. the following equation is added to the model $\theta_{t+1} = \theta_t + n_t$. $n \sim \mathcal{N}(0, 0.5)$ is taken.
- a SIR filter, using 1900 particles, implementing the marginalization method, which consists in getting rid of θ , without trying to estimate it, given the knowledge of $p(\theta)$. s_t and θ are assumed independent, then to update the particles weights in the SIR filter, the relation $p(z_t|s_t) = \int p(z_t|s_t, \theta)p(\theta)d\theta$ is used. $p(\theta)$ the uniform distribution over $[10, 12]$ is taken.



(a) KLD between the reference, and the particle filters using the marginalization method and the artificial dynamics method.



(b) State estimates.



(c) Typical values for θ estimates obtained by the particle filter using the artificial dynamics method.

Figure 8. Example of close distributions discrimination.

These two filters are compared to a reference, namely, a SIR filter PF_{Ref} , with 2000 particles, for which θ is known. The KLD has been computed by $\hat{D}_{KL}^{(10)}$, and only the nearest

(different) neighbor is taken into account in the estimation. The KLD between two runs of the reference is also computed (see figure 8(a)), in order to guarantee that the variance of the estimator is small enough to perform a meaningful comparison.

Simulation results: the blue curve on figure 8(a) indicates us that the variance of the estimator $\hat{D}_{KL}^{(10)}$ is small enough compared to the order of magnitude of the two KLD estimates (represented by the red and the green curve on the same figure), to make their comparison possible. Moreover, correlations between figures 8(a) and 8(c) can be observed: the bigger the KLD, the bigger the error for θ , as well as between figure 8(a) and 8(b): the bigger the state error, the bigger the error for θ . In this simple example, the ability of the KLD to compare close clouds has been demonstrated.

Remark: Another way to reduce the variance of the estimator is, as mentioned in Section II-C, to increase the k_i and l_i in equation (2), if the sample sizes are large enough. Nevertheless, one should be aware that it could introduce a bias for too large values of k_i and l_i . Moreover, there is no general rule to know how much one can gain, since this gain will depend on the dimension of the density and the number of samples used.

E. Interchangeability

One may think that, since drawing once 10^4 samples is identical to drawing 10 times 10^3 samples, one should obtain the same results, with $\hat{D}_{KL}^{(10)}(p^{(10^3)}, q^{(10^3)})$ (p and q are known through 10^3 samples), as with $\hat{D}_{KL}(p^{(10^4)}, q^{(10^4)})$ (p and q are known through 10^4 samples).

Now, in particle filter applications, a large number of particles can be used if the simulation cannot be repeated. On the contrary, a large number of simulation with less particles can be done, if the simulation can be repeated, in order to reduce computational costs⁶.

However, the number of samples drawn for each simulation affects the estimation of the KLD, since the k -nearest neighbors approach is used, of which accuracy depends on the shape and dimension of the estimated density. Consequently, the number of simulations and the number of samples drawn for each simulation might not always be interchangeable, especially when a low number of samples is used.

For the following simulation the KLD between PF_{Ref} and PF_{AD} from the previous section, is computed with $\hat{D}_{KL}^{(n_i)}$, for different n_i and for different N_p , with N_p the number of particles of PF_{Ref} and PF_{AD} . The results are compared with the case $n_i = 1$ and $N_p = 10^4$ to obtain figure (9).

⁶Indeed, the complexity grows linearly with the number of simulations done, but faster than linearly with the number of samples used (see [1]).

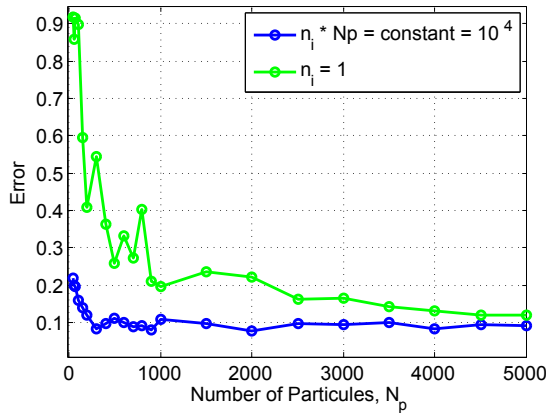


Figure 9. For different N_p , \hat{D}_{KL} and $\hat{D}_{KL}^{(n_i)}$, with $n_i \times N_p = \text{constant} = 10^4$, are compared to \hat{D}_{KL} with $N_p = 10^4$.

Simulation results: we observe in figure (9) that, if N_p is large enough (i.e the density estimate through the k -nearest neighbors approach is good enough), the interchangeability between the number of simulations n_i and N_p might be possible, that is, the same results are obtained when the product $n_i \times N_p$ is fixed, whatever the values for n_i and N_p .

IV. CONCLUSION

A generally applicable comparison tool for particle filters has been presented. This tool is based on the KLD, which performs the comparison of two particle clouds. The estimator of the KLD used has been shown to be practically applicable to any cloud from any particle filter and has the following characteristics:

- small bias and variance for a limited number of particles;
- handles clouds of different size;
- handles clouds with high dimensionality;
- discrimination of close distributions;
- interchangeability, if enough samples are used (i.e the density estimate through the k -NN approach is good enough).

REFERENCES

- [1] M.S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking". *IEEE Signal Processing*, vol. 50, no. 2, pp.174-189, 2002.
- [2] G. Hendeby, R. Karlsson, F. Gustafsson and N. Gordon, "Performance issues in non-Gaussian filtering problems". *Nonlinear Statistical Signal Processing Workshop*, Cambridge, UK, 2006.
- [3] J. Ramírez, J. C. Segura, C. Benítez, A. de la Torre and A. Rubio, "A new Kullback-Leibler VAD for speech recognition in noise". *IEEE Signal Processing*, vol. 11, no. 2, pp.666-669, 2004.
- [4] J. Goldberger, S. Gordon and H. Greenspan, "An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures". *IEEE International Conference on Computer Vision*, vol. 1, pp. 487-493, 2003.
- [5] C. Liu and H.Y. Shum, "Kullback-Leibler boosting". *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2003.
- [6] T. M. Cover and J.A. Thomas. "Elements of information theory", Wiley, 1991.

- [7] D. Fox, "Adapting the Sample Size in Particle Filters Through KLD-Sampling". *International Journal of Robotics Research*, vol. 22, 2003.
- [8] A. T. Ihler, J. W. Fisher III and A. S. Willsky, "Particle filtering under communications constraints". *IEEE Statistical Signal Processing*, 2005.
- [9] L.L. Ong, T. Bailey, H. Durrant-Whyte and B. Upcroft, "Decentralised Particle Filtering for Multiple Target Tracking in Wireless Sensor Networks". *Proceedings of the 11th International Conference on Information Fusion*, pp.1-8, 2008.
- [10] Q. Wang, S.R. Kulkarni and S. Verdú, "Divergence Estimation for Multidimensional Densities Via k-Nearest-Neighbor Distances". *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp.2392-2405, 2009.
- [11] D. W. Scott, "Multivariate Density Estimation". Wiley, 1992.
- [12] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering", in *Sequential Monte Carlo in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. New York: Springer-Verlag, 2001, pp.197-223.